Linking and Integrating two Electronic Lexicons

Nilda Ruimy, Adriana Roventini, Rita Marinelli, Marisa Ulivieri
Istituto di Linguistica Computazionale – CNR
Via Moruzzi,1 – 56124 – Pisa, Italy
{nilda.ruimy,adriana.roventini,rita.marinelli, marisa.ulivieri}
@ilc.cnr.it

Abstract

Nowadays, in the field of Computational Lexicography, much attention is being paid, when building lexical resources, to their interoperability and their easy integration in HLT-NLP applications for an enhanced performance. Concerning already existing computational lexicons, on the other hand, their integration and interoperability is attainable, provided their main features offer a field of comparison. The two largest and extensively encoded electronic lexicons of Italian language fulfill this essential requirement. Although developed according to two different lexical models, ItalWordNet and PAROLE-SIMPLE-CLIPS present in fact many compatible aspects. Linking and eventually merging these lexical resources in a common representation framework seems therefore a wise move to offer the end-user a more exhaustive and in-depth lexical information combining the potentialities and most outstanding features offered by the two lexical models. This paper reports on the ongoing linking of the two lexicons. The mapping of the ontologies on which basis the lexicons are structured is described; an overview of the adopted methodology, of the linking process and of the results of the first mapping phase regarding 1st Order Entities is provided. Reciprocal benefits and enhancements for the two resources are also illustrated that definitely justify the soundness of our linking initiative.

1 Introduction

'As the need for cross-lingual studies and applications grows, it is increasingly important to develop resources in the world's languages that can be compared and linked, used and analyzed with common software, and that contain linguistic information for the same or comparable phenomena.'[1].

Nowadays, in the field of Computational Lexicography, much attention is being paid, when building lexical resources, to their interoperability and their easy integration in HLT-NLP applications for an enhanced performance. The most relevant collaborative efforts that lexicon experts devoted to developing consensual specifications and enforcing standards in this domain have led to the creation of the Lexical Markup Framework (LMF)[2], a metamodel which provides a common standardized framework for the construction of computational lexicons.

Concerning already existing electronic lexicons, on the other hand, their integration and interoperability is attainable, provided their main features offer points of comparison. The two largest and extensively encoded lexicons of Italian language, which were developed during the last decade at the CNR Institute of Computational Linguistics in Pisa, fulfill this essential requirement.

[1] First International Conference on Global Interoperability for Language Resources ICGL2008, Call for Papers, Mission.
[2] in the International Organization for Standardization (ISO) sub-group TC37/SC4/WG4.

2 Lexical Resources

ItalWordNet[3] (henceforth IWN) is a lexical semantic database created in the framework of the *EuroWordNet* (EWN) project[4] and extended in the national project *Integrated System for the Automatic Language Treatment* (SI-TAL). It is based on the EuroWordNet lexical model[5] (Vossen, 1998) which is, in turn, inspired to the Princeton WordNet (Miller *et al.,* 1990).

IWN (Roventini *et al*., 2003) provides the semantic description of 67,000 Italian word senses (verbs, common and proper nouns, adjectives, adverbs and multi-word units), which are clustered in about 50,000 *synsets* (i.e. synonym sets). One of the salient features of the resource is the connection of all IWN synsets to the Princeton Wordnet database (Fellbaum, 1998). Such synsets, that represent lexicalized concepts, are classified in terms of an ontology and interconnected by means of a set of semantic relations that link both intracategorial and intercategorial synsets (Alonge *et al.* 1998).

The IWN Top Ontology (henceforth, TO), which slightly differs from the EWN TO[6], is a hierarchy of 65 language-independent Top Concepts (TCs) clustered in three main categories distinguishing 1stOrderEntities, 2ndOrderEntities and 3rdOrder Entities. Their subclasses, hierarchically ordered by means of a subsumption relation, are also structured in terms of (disjunctive and non-disjunctive) opposition relations.

PAROLE-SIMPLE-CLIPS[7] (henceforth PSC[8]) is a four-layered lexicon developed over three different projects. Morphological and syntactic models and the kernel of related lexicons were elaborated in the EU *LE-PAROLE* project; the semantic model and the core of the semantic lexicon, in the EU *LE-SIMPLE* project[9]; the phonological level of description and the extension of the lexical coverage were performed in the context of the Italian national project *Corpora e Lessici dell'Italiano Parlato e Scritto* (CLIPS). The theoretical model underlying this lexicon is based on the EAGLES recommendations, on the results of the EWN and ACQUILEX projects and on a revised version of Pustejovsky's Generative Lexicon theory (Pustejovsky 1995).

At the semantic level, the PSC lexicon (Ruimy *et al.* 2003), which comprises more than 57,000 Italian word senses (verbs, common and proper nouns, adjectives, adverbs and grammatical words), is structured in terms of an ontology.

The SIMPLE Ontology[10] (SO) consists of 157 language-independent semantic types designed for the multilingual lexical encoding of concrete and abstract entities, properties and events. It is a multidimensional type system, based on hierarchical and non-hierarchical conceptual relations, which distinguishes between *simple* (one-dimensional) and *unified* (multi-dimensional) semantic types, the latter implementing the principle of *orthogonal inheritance* (Pustejovsky & Boguraev, 1993). Multidimensionality is captured by *qualia roles* that define the distinctive properties of semantic types and differentiate their internal semantic constituency.

Since IWN, unlike PSC, is a one-layer lexical database, the comparison of the resources focuses on their semantic information. In this regard, each lexicon provides a bundle of specific properties reflecting the different principles and peculiarities that characterize its underlying model[11] but also a large number of conceptually similar information that represent the compatible aspects of these two lexicons. In this connection, it is worth reminding that EWN was one of the inspiration sources for the SIMPLE model of semantic representation.

Studying the two resources, the wide range of compatibility observed did prompt us to undertake their semi-automatic link, eventually combining and merging the whole information into a common representation framework. In this respect, LMF, which enables the merging of electronic lexical resources, seems an appropriate candidate framework all the more since its creation was largely inspired by the PAROLE-SIMPLE model.

The remainder of this paper reports on the mapping of the ontologies on which basis both

[3] http://www.ilc.cnr.it/viewpage.php/sez=ricerca/id=834/vers=ing
[4] http://www.hum.uva.nl/~EWN
[5] The only aspects in which IWN differs from EWN are a few amendments made to the ontology in order to allow for the representation of adjectives and the addition of further lexical-semantic relations.
[6] Cf. note5.
[7] http://www.ilc.cnr.it/clips/CLIPS_ENGLISH.htm
[8] 'PSC' is not the acronym of the lexicon. It is only used here for brevity
[9] http://www.ub.es/gilcub/SIMPLE/simple.html
[10] http://www.ilc.cnr.it/clips/Ontology.htm
[11] such as, for example, a different ontological framework and a different approach to the organization of lexical units

lexicons are structured; it provides an overview of the linking methodology developed and of the mapping process implemented in an Access software tool; it illustrates the results of the first mapping phase, which was devoted to 1stOrderEntities. Resulting from the mapping, some reciprocal benefits for the resources are also illustrated that definitely justify the soundness of our linking initiative. Ongoing and future work is outlined in the conclusion.

3 Mapping the Ontologies

3.1 Ontological typing

Let us first very briefly illustrate the ontological typing in the two models.

According to the SIMPLE model, the basic unit, i.e. the word sense, represented by a 'semantic unit' (*SemU*) is associated to *one single* semantic type, e.g.: SemU69589*cardiologia* (cardiology) [DOMAIN]; USem4985*insegnante* (teacher): [PROFESSION]. Through its type membership, each *SemU* is endowed with a structured set of semantic features and relations; among these are the 60 relations of the *Extended Qualia structure*, a revisited version of the original GL representational tool that enables to describe both the componential aspect of a word meaning and its relationships to other lexical items.

The EWN/IWN model, by contrast, allows for a multi-classification. Synsets are in fact seldom linked to one single ontological node but rather cross-classified in terms of multiple, non-disjoint, TCs[12], e.g.: synset29146: {N *cardiologia*1}: [Agentive Purpose Social Undboundedevent]; synset4283: {N *docente*1, *didatta*1, *professore*2, *insegnante*1}: [Human Object Occupation]. Noteworthy here is that the ontological classification is determined by the choice of the synset hyperonym. As to the word sense or, in WN parlance, *synset variant,* its semantics is fully defined by its membership in a synset.

Although moving from a different approach to the definition of word sense, the information provided by these two types of ontological classification is substantially equivalent. Owing to the multidimensional nature of the ontology, SIMPLE types encompass in fact the various meaning dimensions that are expressed in IWN by the different TCs cross-classifying 1st and 2ndOrder Entities.

3.2 Mapping the ontological classes

In the process of mapping these two ontology-based lexical resources, the first step clearly consisted in comparing their ontological framework, viz. in manually establishing correspondences between the conceptual classes of both ontologies, with a view to further matching their respective instances. The comparison was done so far for classes structuring entities and events (Ruimy, 2005)[13]; the ontological typing of adjectives will be dealt with in a further phase of work.

A preliminary observation to be done is that IWN TO consists of a set of rather flat top semantic features whereas SO encompasses mono- and multi-dimensional types with associated templates of structured information that define the content of the conceptual types.

As mentioned in section 2, the first subdivision level of IWN TO consists of three main classes:

The 1stOrderEntity class structures concrete entities (referred to by concrete nouns). Its main cross-classifying subclasses: Form, Origin, Composition and Function correspond to the four Qualia roles the SIMPLE model avails of to express orthogonal aspects of word meaning. Their respective subdivisions consist of (mainly) disjoint classes, e.g. 'Natural' vs. 'Artifact', 'Substance' vs. 'Object'. To each class corresponds, in most of the cases, a SIMPLE semantic type or a type hierarchy subsumed by the CONCRETE_ENTITY top type. Some other TCs, such as 'Comestible' and 'Liquid', are instead mappable to SIMPLE distinctive features Plus_Edible, Plus_Liquid, etc.

The 2ndOrderEntity class classifies static or dynamic situations denoted by nouns and verbs, adjectives and adverbs. 2ndOrderEntities are primarily characterized in terms of two classification parameters: 'Situation Type' – whose two disjoint features, 'Static' and 'Dynamic' encode the event structure – and 'Situation Component', which subsumes a set of combinatorial classes providing a more conceptual classification in terms of semantic components of a

[12] The more specific the word, the more TCs contributing to its description.

[13] http://www.ilc.cnr.it/clips/Ontology_mapping.doc

concept, e.g.: ‘Manner’, ‘Experience’, ‘Communication’, ‘Cause’.

Concerning the Situation Type, in the SIMPLE model, the event structure is expressed by means of the three-valued feature Eventtype = *state*, *process*, *transition*, values which correspond in IWN respectively to ‘Static’, (Dynamic) ‘Unbounded-event’ and (Dynamic) ‘Boundedevent’. As to the combinatorial subclasses of the Situation Component, each one generally corresponds to one or more SIMPLE types, depending on the Situation Type value and/or the other Situation Components it combines with, as illustrated in table 1.

IWN Top Concepts	SIMPLE semantic type
Existence Bounded Cause Physical	CREATION, CAUSE_NATURAL_TRANSITION
Existence Static	EXIST
Experience Mental Dynamic	EXPERIENCE_EVENT, MODAL_EVENT
Experience Physical Stimulating Dynamic	PERCEPTION

Table 1. TCs combinations and semantic type correspondences

The 3rdOrderEntity class, which has no further subdivision, classifies abstract entities (denoted by abstract nouns) existing independently of time and space. These entities fall into the ABSTRACT_ENTITY type hierarchy of the SIMPLE ontology.

Notwithstanding the different approaches taken for their design and some different underlying principles, these two ontologies globally show a significant degree of overlapping and no fundamental difference in conceptualization is observed. Two general remarks are in order here:

1) Owing to the different extension of both ontologies, some specific concepts – which are expressed in SIMPLE Ontology by lower level semantic types – are likely to have no equivalent in IWN TO.

2) Not surprisingly, mapping from event-denoting PSC semantic units to IWN 2ndOrderEntities immediately appears more challenging than dealing with 1rstOrderEntities that pose less tricky problems.

4 Linking Methodology

Owing to a different organizational structure of information in the two resources, the linking process involves elements having a different status, viz. autonomous semantic units in PSC and synsets clustering 1 to n synset variants in IWN.

In order to avoid dealing with huge, unmanageable sets of data, mapping is performed on a semantic type-driven basis and is PSC → IWN oriented. The rationale for this orientation is that the 157 semantic types of the SO provide a more fine-grained structure of the lexicon than the 65 top concepts of the IWN ontology, which reflect only fundamental distinctions.

Taking therefore as starting point the lexical instances of a SIMPLE semantic type along with their PoS and hyperonymic (‘isa’) information, the IWN resource is explored in search of linking candidates.

Each mapping run returns two data sets:

► Matched pairs of word senses, i.e. SemUs and synset variants with identical string and PoS and whose ontological classification matches the correspondences established between the classes of both ontologies.

These word senses are linked after human validation.

Linking may occur between a *SemU* and a *one-variant* synset (1):

1. SemU66448bastone ↔ synset29146 {N *bastone*1}

or between a *SemU* and one word sense of a *multi-variant* synset (2):

2. SemU75412*adornare* ↔ synset35336 {V *adornare*1, *ornare*1, *decorare*1, *guarnire*3, *addobbare*4}

► Unmatched word senses, in spite of their identical string and PoS value. Matching failure may be due either to coverage discrepancies (lack, in IWN, of a lexical item or of the appropriate word sense corresponding to a PSC entry) or to a mismatch of ontological classification between word senses existing in both resources. Focusing on this latter case, two main obstacles hamper their matching:

1) An incomplete ontological information:

As already said, IWN synsets are cross-classified in terms of a combination of TCs. This combined notation is however sometimes only partially encoded and cases are not rare of 1rstOrderEntities lacking some meaning component or

2nd OrderEntities lacking one of their two classifying parameters.

For the linking purpose, the problem of incomplete ontological classification may, in a number of cases, be overcome by relaxing the mapping constraints. Yet, this solution can only be applied if the existing ontological classification, in spite of its incompleteness, is informative enough. More problematic to deal with are those cases of incomplete and little informative ontological labels. This is the case, for example, of 1rst OrderEntities as different as *medicinale, anello, laccio, vetrata* (medicine, ring, lace, glass window) and only classified as 'Function' or of 2nd OrderEntities lacking either a Situation Component or a SituationType, e.g. *unirsi* (to join) classified as 'BoundedEvent' or *sciogliere* (to melt) as 'Cause'.

2) A different ontological information:

Besides mere encoding errors for which a correction phase is foreseen, the ontological classification may be different with respect to the constraints imposed to the mapping run and the discrepancy may be imputable to:

i) A different but equally defensible meaning interpretation in each resource, e.g.: *ala* (aircraft wing): 'Part' vs. 'Artifact Instrument Object'. Word senses falling into this category are clustered into numerically significant sets according to their semantic typing and then studied with a view to establishing further equivalences between ontological classes or to identify, in their classification schemes, descriptive elements lending themselves to comparison.

ii) The presence of polysemous senses of the considered *SemUs* (e.g., USem65931*kiwi* 'Fruit' which is obviously discarded when mapping the *kiwi* instance of the 'Animal' class). Some of these word senses proceed from an extension of meaning, e.g. People-Human: *pigmeo, troglodita* (pygmy, troglodyte) or Animal-Human *verme, leone* (worm, lion) and are used with different levels of intentionality: either as a semantic surplus or as dead metaphors (Marinelli, 2006).

iii) A different level of specificity in the ontological classification, either due to the lexicographer's subjectivity or to an objective difference of granularity of the ontologies, cf. the *viola* example below.

Problems emerging with instances of iii) may be bypassed by climbing up the ontological hierarchy, identifying the parent nodes and allowing them to be taken into account in the mapping process.

Hyperonyms of matching candidates are also consulted during the mapping process and play a particularly determinant role in the resolution of cases whereby matching fails due to a conflict of ontological classification, namely:

● sets of word senses displaying a different ontological classification in each resource but sharing the same hyperonym, e.g. *collana, orecchino* (necklace, earring) are typed as CLOTHING in PSC and as 'Function' in IWN but share the hyperonym *gioiello* (jewel).

● polysemous senses belonging to different semantic types in PSC but sharing the same ontological classification in IWN, e.g.: in PSC, SemU1595*viola* (violet) PLANT and SemU1596*viola* FLOWER vs. in IWN: *viola*1 (has_hyperonym *pianta*1) and *viola*3 (has_ hyperonym *fiore*1) (flower), both typed as 'Group Plant'.

5 Mapping Process

The Access software tool devised to map the lexical units of both lexicons works in a semi-automatic way using the ontological classifications, the hyperonymic relations and some semantic features of the two resources. The mapping process foresees the following steps:

► Selection of a PSC semantic type and definition of the search range, i.e. either all of its instances or a subset bearing a selected feature, e.g. PLANT and 'Plus_Edible';

► Selection of one or more mapping constraints on the basis of the correspondences established between the conceptual classes of both ontologies;

► Human validation of the automatic mapping, i.e. selection of the semantically relevant word sense pair(s) from the set of possible matches automatically output for each *SemU* (referred to as *multiple mapping* in table 2). Multiple mappings depend on the more fine-grained sense distinctions performed in IWN. Cases are in fact frequent of a single entry in PSC corresponding to two different IWN entries encoding very fined-grained nuances of sense, e.g.: SemU63617*galeotto* vs. synset28576: {N *galeotto*1} (galley rower) and synset49579: {N *galeotto*2} (galley slave). The selection of the relevant word sense pair involves checking information sources such as hyperonyms, SemU / synset glosses and ILI links;

► Relaxation / tuning or addition of mapping constraints, where appropriate; new processing of the input data.

6 Mapping Results

The results of the first working phase, which was devoted to linking concrete entities, sound quite encouraging since 72,32% of the word senses considered have been successfully linked. Table 2 evidences: i) the extent of overlapping coverage for concrete entities; ii) the considerable percentage of linked senses with respect to the linkable ones (i.e. words with identical string and PoS value); iii) the many cases of multiple mappings.

Overlapping coverage		56,29%
Selected *SemUs*	27,768	--
Linkable senses	15,193	54,71%
Linked senses	10,988	72,32%
Multiple mappings	1,125	10,23%
Unmatched senses	4,205	27,67%

Table 2. Mapping concrete entities

7 Enhancement of the Resources

Besides offering the end user a more exhaustive and in-depth lexical information combining the potentialities and most outstanding features of the two lexical models, the linking process lets inconsistencies that unavoidably exist in both resources emerge, allowing therefore to amend them. To give but an example, consistency would require that, when a *synset variant* is linked to a *SemU*, all the other variants from the same synset map to PSC entries sharing the same semantic type. Yet, especially concerning event denoting words, cases have been observed whereby *SemUs* corresponding to variants of the same synset do not share a common SIMPLE semantic type. Linking the two resources permits therefore to enhance their consistency since it implies a de facto reciprocal assessment of both coverage and accuracy, which is particularly relevant to hand-built lexical resources. ‘Cleaning’ the two lexical resources represents moreover a step forward towards their interoperability and eases therefore their eventual merging.

Moreover, the linking process makes it possible to enrich each resource by complementary information types that are peculiar to the other’s theoretical model. In EWN, the richness of sense distinctions and the consistency of hierarchical links are remarkable. SIMPLE, on the other hand, focuses on richly describing the meaning and semantic context of a word and on linking its syntactic and semantic representation, which is crucial for most NLP applications.

7.1 IWN Information Enriching PSC SemUs

The organization of lexical knowledge has entailed a quite coherent clustering of synonyms in Wordnet-based resources. The SIMPLE model, on the other hand, has devoted more attention to other relation types and less importance has been given to the instantiation of synonymy links.

Integrating the two lexicons, PSC entries will easily be enriched by synonymy links, based on synset membership. Likewise, missing senses of existing words and new lemmas will be quickly and consistently encoded in PSC lexicon.

In IWN, hierarchical links are of fundamental importance and hence consistently expressed by two relations ‘has_hyperonym’ and ‘has_hyponym’ that allow a cross-checking of data. In the SIMPLE model, on the other hand, the focus put on covering the whole range of a word’s syntactic and semantic uses has sometimes prejudiced the enforcement, in PSC entries, of coherent taxonomic links and yielded cases of circularity. Such cases will be amended by resorting to IWN hyperonymic links for nouns and verbs.

IWN ‘Involved_agent / patient / instrument / location’ and ‘Role_agent / patient / instrument / location’ semantic relations, respectively linking 2^{nd} with 1^{rst} or 3^{rd}OrderEntities and conversely will be most helpful to relate more straightforwardly, in PSC lexicon, an event to its participants[14], e.g.: *operare*: involved_agent = *chirurgo*, involved_ patient = *paziente*, involved_location = *ospedale*, involved_instrument = *bisturi* (to operate, surgeon, patient, hospital, lancet); an entity to an event: *studente*: role_patient = *insegnare* (student, to teach) or even to relate event’s participants to each other: *chirurgo*: co_agent_patient = *paziente*. These links will moreover allow to discriminate the nature of some relationships that are rather poorly rendered, in the PSC lexicon, by the overused — and hence misused — constitutive relation ‘concerns’, e.g.: 1)

[14] Work on this issue is now in progress (Ruimy, 2007).

otturazione concerns *dente* (filling, tooth)*;* 2) *sbarcare* concerns *nave* (disembark, boat) could be respectively replaced by involved_patient and **i**nvolved_source_direction.

7.2 PSC Information Enriching IWN Synsets

No argument structure information is provided in IWN. Linking the two lexicons, IWN predicative units will be endowed with information concerning their syntactic and semantic subcategorization frames.

IWN word senses will also inherit the PSC extensively encoded information concerning their domain of use. Such information, most relevant to IR, WSD, IE and parsing, enables – among other – clustering semantically related lexical items pertaining to specific domains, regardless of their PoS and type membership.

Given the rich lexical information foreseen by the SIMPLE model, IWN synsets will also gain:
- a finer-grained ontological classification: let us observe for example that, as against the 'Plant', 'Human', and 'Communication' TCs, SIMPLE Ontology offers respectively 5, 9 and 8 semantic types, each one providing a rich bundle of specific information;
- a semantic description less prominently based on taxonomic relations. SIMPLE semantic relations, which are defined along multiple dimensions, enable to avoid an overloading of the 'isa' relation and to represent senses not adequately definable in terms of the hyperonymic link.
- the expression of further orthogonal meaning dimensions: e.g., synset variants such as *inchiostro* or *colla* (ink, glue), associated to the TC 'Substance' and bearing telic information, will acquire, through their linking to the corresponding PSC entries, constitutive and agentive dimensions.
- the account of systematic polysemy. Regular polysemy is expressed, in the SIMPLE model, through distinct entries connected by means of a polysemous relation linking the ontological typing of pairs of senses, according to a set of polysemous classes, e.g. *banca* (bank): BUILDING / INSTITUTION; LOCATION / HUMAN_GROUP. In IWN, such polysemous senses are encoded as separate meanings but no mention is made about the way they relate to each other. The possibility provided by the EWN model to assign two or three disjunctive hyperonyms was in fact not systematically implemented. During the IWN project, a proposal was made to encode regular polysemy using the 'is_extension_of / has_extension' relation, originally created for proper nouns (Marinelli, 2004).
- a more specific identification of the nature of some syntagmatic relationships not expressible in the IWN model, and which are, for instance, most relevant for extracting semantic networks, e.g.: *antidoto* 'used_against' *veleno* (antidote, poison), *acetone* 'used_as' *solvente* (acetone, solvent).

8 Conclusion and Future Work

This paper reported on the ongoing linking of the two largest general-purpose, electronic lexical resources of Italian language, PAROLE-SIMPLE-CLIPS and ItalWordNet. The mapping of the ontologies on which basis the lexicons are structured was described. An overview of the adopted methodology, of the linking process and of the results of the first mapping phase regarding 1stOrderEntities was provided. Reciprocal benefits and enhancements for the two resources were also illustrated.

Differences regarding the nature of linking units, the granularity of sense distinction and the ontological typing are complex issues that are being addressed during the linking process. Problems arise, in particular, when encoding incompleteness or inconsistency generate unpredictable, non-systematic ontological typing discrepancies whereby a theoretical comparison of the models evidenced a high degree of overlapping. Nevertheless, the wide range of compatibility the models show induces us to believe that semantic interoperability is indeed achievable and it is our strong conviction that linking two resources based on such valuable and widely tested lexical models that have addressed challenging (and complementary) research issues in lexical semantics is a most appropriate and timely initiative. Semantic integration of these resources is all the more desirable considering their multilingual vocation: IWN is linked to the WN of seven other languages and PSC shares with eleven European lexicons a theoretical model, representation language, building methodology and a core of entries.

On the basis of the encouraging results obtained from the linking of 1rstOrderEntities, work is now in progress as regards the mapping of

3rdOrderEntities and 2ndOrderEntities which, so far, had only been object of preliminary investigations on Speech act (Roventini, 2006) and Feeling verbs.

References

Adriana Roventini, Marisa Ulivieri and Nicoletta Calzolari. 2002. *Integrating two semantic lexicons, SIMPLE and ItalWordNet: what can we gain?* LREC 2002, Third International Conference on Language Resources and Evaluation Proceedings, Vol. V, pp. 1473-1477, Las Palmas de Gran Canaria.

Adriana Roventini et al. 2003. ItalWordNet: *Building a Large Semantic Database for the Automatic Treatment of Italian*. Computational Linguistics in Pisa, Special Issue, XVIII-XIX, Pisa-Roma, IEPI, vol. II, 745-791.

Adriana Roventini and Nilda Ruimy 2005. *Linking and harmonizing different lexical resources: a comparison of verbal entries in ItalWordNet and PAROLE-SIMPLE-CLIPS.* In Proceedings of the Third International WordNet Conference, Jeju Island, Korea

Adriana Roventini. 2006. *Linking Verbal Entries of Different Lexical Resources*. LREC Proceedings, CD-ROM, 1710-1715.

Alessandro Lenci et al. 2000. *SIMPLE Linguistic Specifications*, Deliverable D2.1, ILC-CNR, Pisa.

Antonietta Alonge, et al. 1998. *The Linguistic Design of the EuroWordNet Database*, Special Issue on EuroWordNet. In N. Ide, D. Greenstein, P. Vossen (eds.), 'Computers and the Humanities', XXXII, 2-3, (91--115).

Christiane Fellbaum (ed.) 1998. *Wordnet: An Electronic Lexical Database*. MIT Press.

George Miller et al. 1990. *Introduction to WordNet: An On-line Lexical Database*, International Journal of Lexicography, III, 4, 235--244.

Gil Francopoulo, et al. 2007. *Lexical Markup Framework: an ISO Standard for Semantic Information in NLP Lexicons*, Proceedings of the Workshop on Lexical-Semantic and Ontological Resources of the GLDV Working Group on Lexicography at the Biennal Spring Conference at the GLDV, Tübingen, (13-14/04/2007).

James Pustejovsky and Branimir Boguraev 1993. *Lexical knowledge representation and natural language processing,* Artificial Intelligence, Volume 63, Special volume on natural language processing, Issue 1-2, 193 – 223.

James Pustejovsky 1995. *The generative lexicon*. The MIT Press, Cambridge, MA.

James Pustejovsky 1998. *Specification of a Top Concept Lattice*, ms., Brandeis University.

Nilda Ruimy et al. 2002. CLIPS, *a Multi-level Italian Computational Lexicon*, LREC 2002, Third International Conference on Language Resources and Evaluation Proceedings, Vol. III, (pp.792-799), Las Palmas de Gran Canaria.

Nilda Ruimy et al. 2003. *A computational semantic lexicon of Italian: SIMPLE*. In A. Zampolli, N. Calzolari, L. Cignoni, (eds.), Computational Linguistics in Pisa, Special Issue, XVIII-XIX, (2003). Pisa-Roma, IEPI. Tomo II, 821-864.

Nilda Ruimy and Adriana Roventini. 2005. *Towards the linking of two electronic lexical databases of Italian*, In Z. Vetulani (ed.), L&T'05, April 21-23, Poznan, Poland. Wydawnictwo Poznanskie Sp. z o.o. 230-234.

Nilda Ruimy 2006. *Merging two Ontology-based Lexical Resources*. LREC Proceedings, CD-ROM, 1716-1721.

Nilda Ruimy 2007. *Enhancing SIMPLE Semantic Relations: A Proposal.* In Zygmunt Vetulani (ed.), Proceedings of 3rd Language & Technology Conference. Fundacja Uniwersytetu im A. Mickiewicza, Poznań. 119-123.

Nuno Silva et al. 2005. *An approach to ontology mapping negotiation*, Proceedings of the Third International Conference on Knowledge Capture Workshop on Integrating Ontologies; Banff, Canada.

Piek Vossen (ed.) 1998. *EuroWordNet: A multilingual database with lexical semantic networks*. Kluwer Academic Publishers.

Piek Vossen (ed.) 2002. *EuroWordNet: General Documentation*, Version 3, Final, July 1.

Rita Marinelli. 2004. *Proper Names and Polysemy: From a Lexicographic Experience*. LREC 2004: Fourth International Conference on Language Resources and Evaluation, Lisbon, Portugal, Proceedings, Volume I, 157-160.

Rita Marinelli. 2006. *Computational Resources and Electronic Corpora in Metaphors Evaluation*. Second International Conference of the German Cognitive Linguistics Association, Munich, 5-7 October.